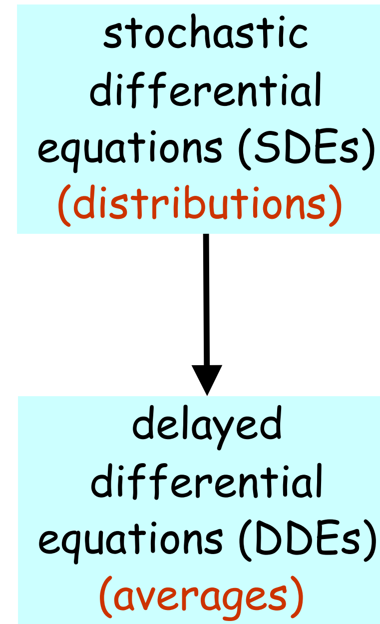
Fluid Models for Large Heterogeneous Networks

W. Gong, C. Hollot, J. Kurose, V. Misra, D. Towsley

http:gaia.cs.umass.edu/fluid/

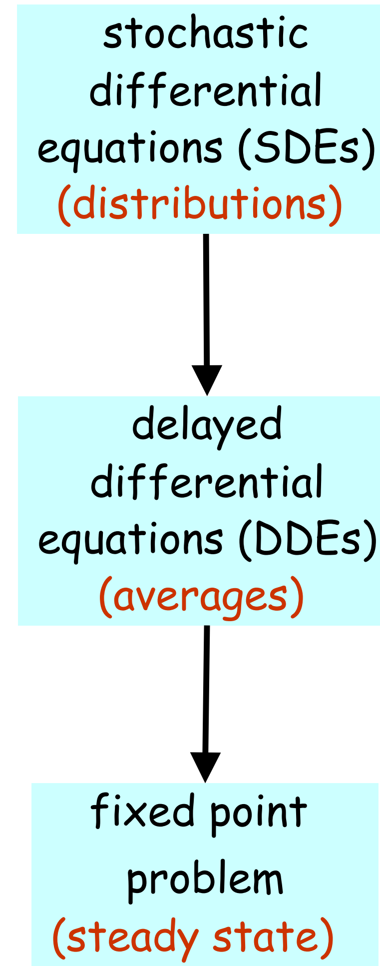
Project goals

- efficient algorithms for transient analysis of large IP networks
  - distributions (SDEs)
  - averages (DDEs)



Project goals

- efficient algorithms for transient analysis of large IP networks
  - distributions (SDEs)
  - averages (DDEs)
- fast algorithms for prediction of steady-state behavior of large IP networks

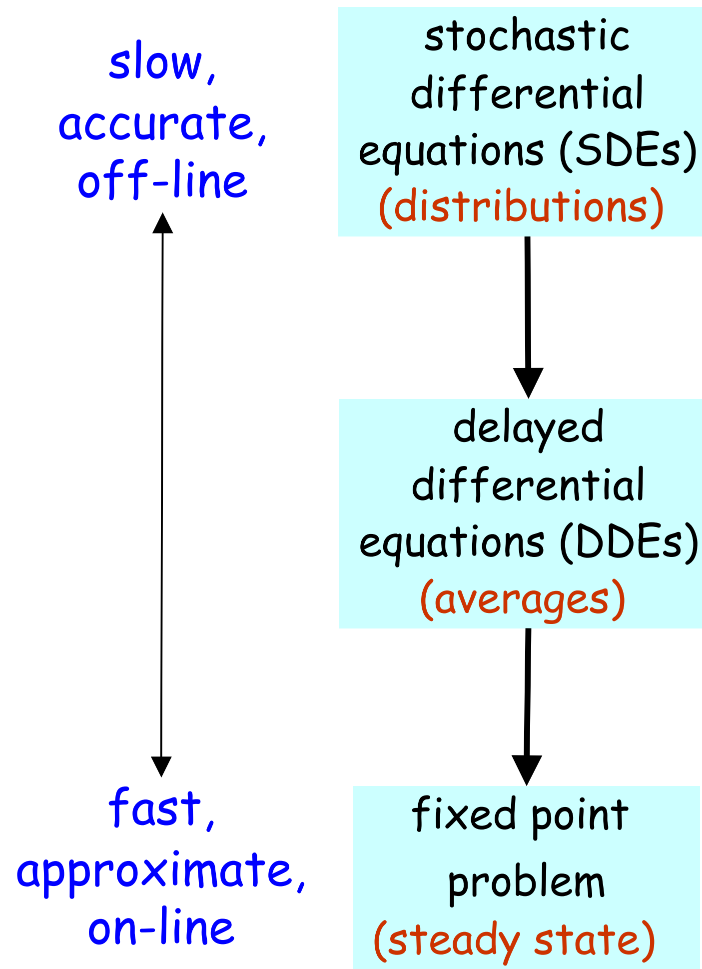


Project goals

- efficient algorithms for transient analysis of large IP networks
  - distributions (SDEs)
  - averages (DDEs)
- fast algorithms for prediction of steady-state behavior of large IP networks

using fluid models

- develop/refine network control algorithms



Modeling results

SDEs - time-stepped techniques

speedup vs. accuracy

DDEs - extensions to DiffServ

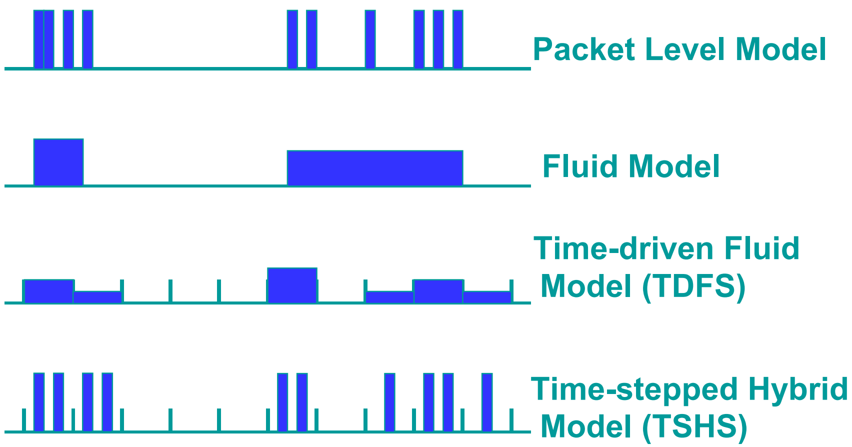
handling QoS

FP - ACK losses, drop tail

greater generality

Time stepped fluid simulation

- divide traffic into fixed length segments
  - segment -> fluid chunk
  - packet info. in fluid chunk
- accurate at high loads
- less accurate at low load, bursts at fine time-scales



To do:

- formal error analysis
- multi-resolution modeling for large, heterogeneous networks.

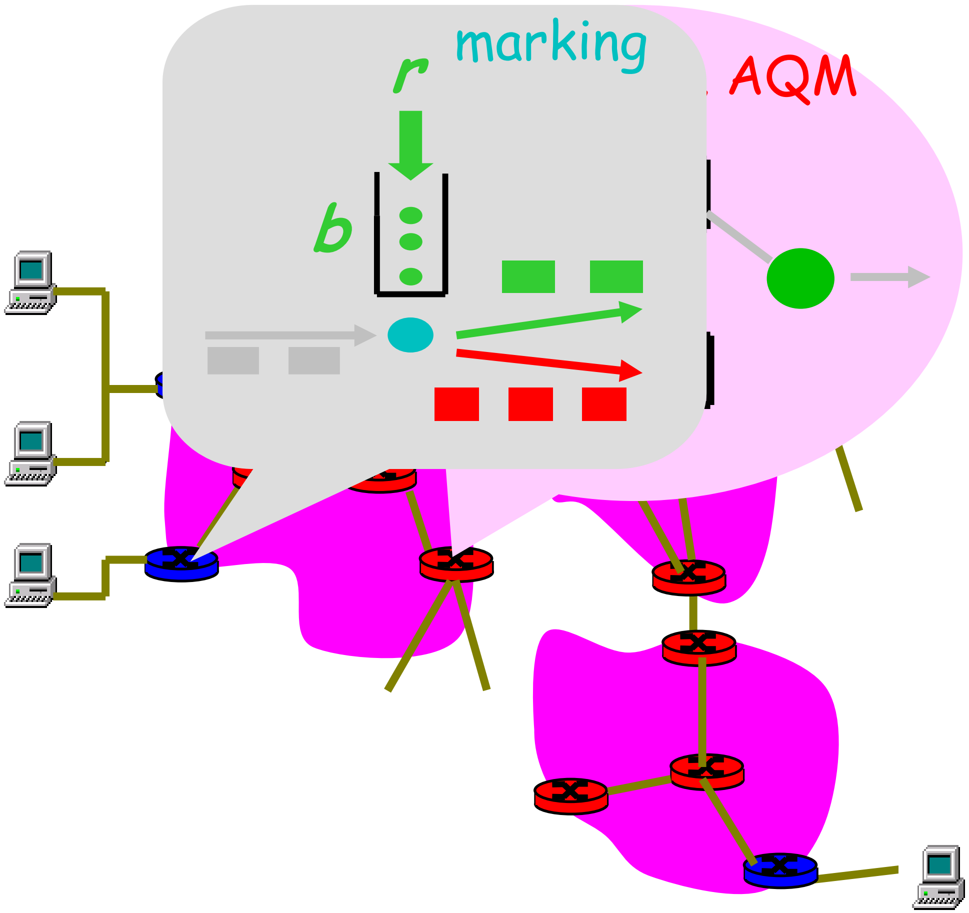
DiffServ architecture

Edge router:

- aggregate traffic management
- marks packets as in-profile and out-profile

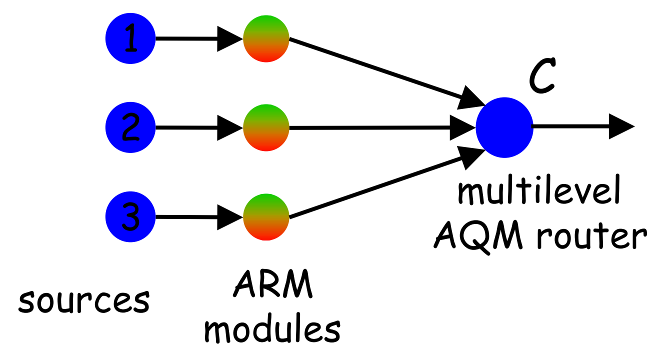
Core router:

- per class traffic management
- buffering and scheduling based on marking at edge
- preference given to in-profile packets
- Assured Forwarding

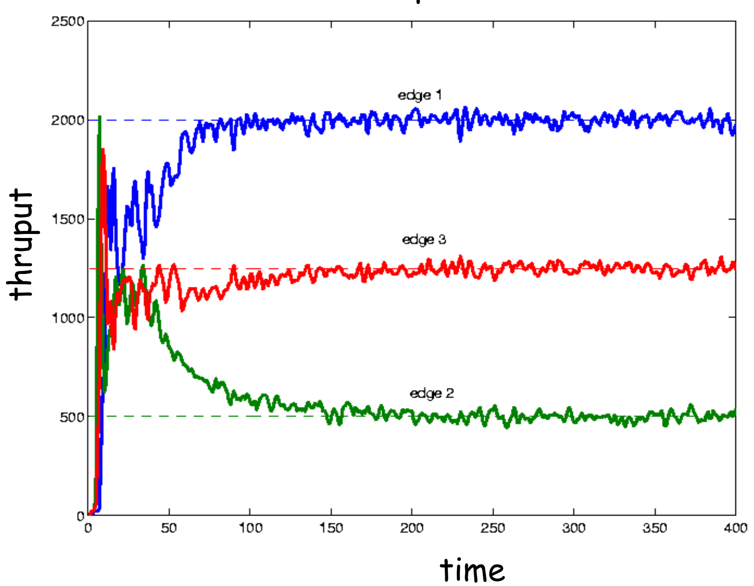


Bandwidth guarantees

- M aggregates, edge markers, target rates $\{A_i\}$
- single bottleneck, capacity C
- adaptive rate management (ARM) at edges
  - monitor achieved thruput
  - PI control to adapt r_i
- multilevel PI control at routers
- SDEs, DDEs describe behavior
- target rates $\{A_i\}$ achievable if $\Sigma A_i \leq C$

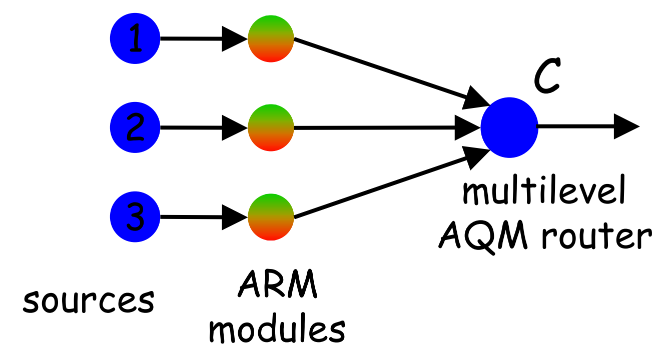


$\Sigma A_i = C$

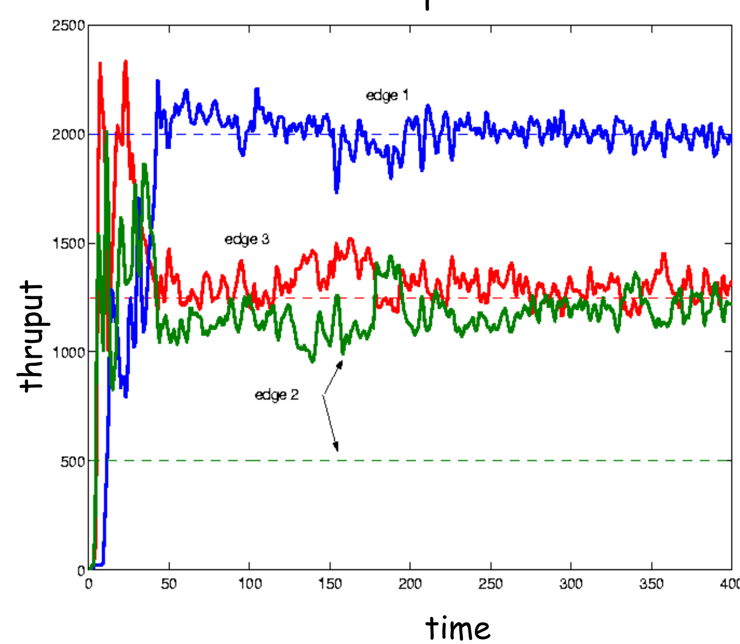


Bandwidth guarantees: solution

- M aggregates, edge markers, target rates $\{A_i\}$
- single bottleneck, capacity C
- adaptive rate management (ARM) at edges
  - monitor achieved thruput
  - PI control to adapt r_i
- multilevel PI control at routers
- SDEs, DDEs describe behavior
- target rates $\{A_i\}$ achievable if $\Sigma A_i \leq C$



$\Sigma A_i = C$



Concurrent downloads

- concurrent download software widely available
  - FlashGet, Go!Zilla, ReGet, Download Accelerator, GetRight, GetSmart, Download Devil
- multiple TCP flows for same object
- analysis shows very aggressive bandwidth usage
  - inherent unfairness
  - prisoner's dilemma
  - network, server congestion
- need to provide servers incentive to cooperate with network

Traffic behavior

- network traffic exhibits correlations over multiple timescales (Leland,...; Floyd, Paxson; ...)
- explanations
  - heavy-tailed web object sizes (Crovella, Bestavros)
  - TCP protocol behavior (Veres, Boda; Feng, etal.;Sikdar, Vastola; Guo, etal.)
- understanding can lead to better network/protocol design

Web object size distribution

- disagreement on tail of web file size distr. (BC97, Downey01)

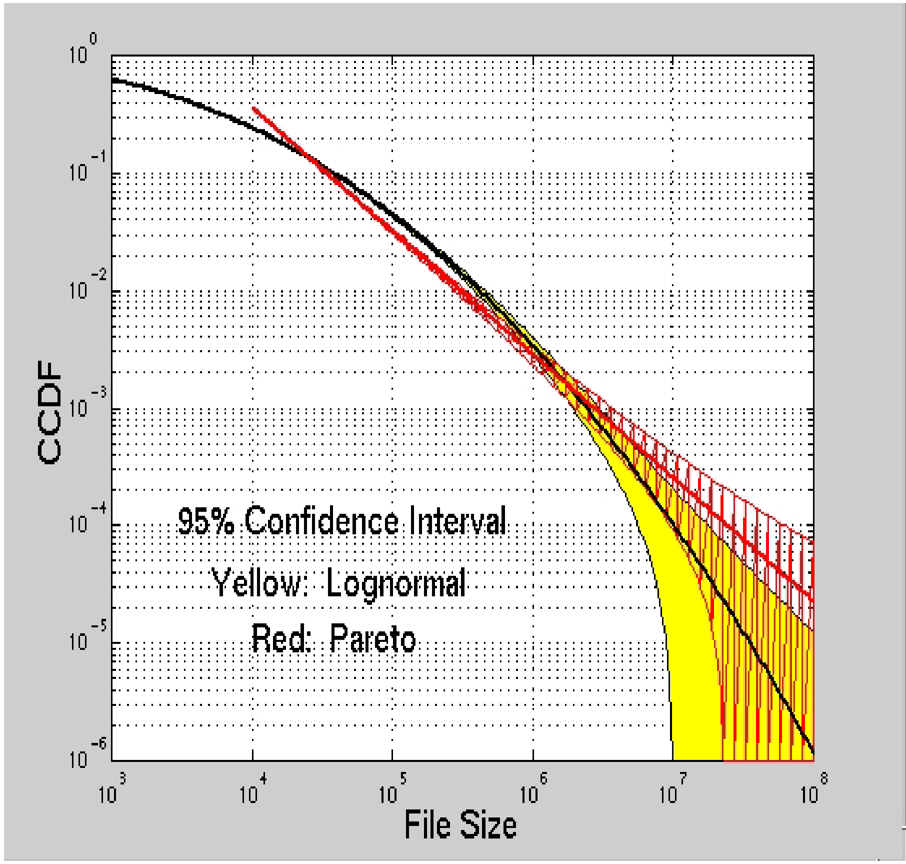
Web object size distribution

- disagreement on tail of web file size distr. (BC97, Downey01)
- competing models agree on body, … but not tail
  - pareto (GBM, HOT, …)
  - lognormal (CLT, …)

Web object size distribution

- disagreement on tail of web file size distr. (BC97, Downey01)
- competing models agree on body, …
  but not tail
  - pareto (GBM, HOT, …)
  - lognormal (CLT, …)
- tails fragile
  - sensitive to perturbation in model assumptions
- finite data inadequate to identify tail
- tails don't affect network engineering, body does

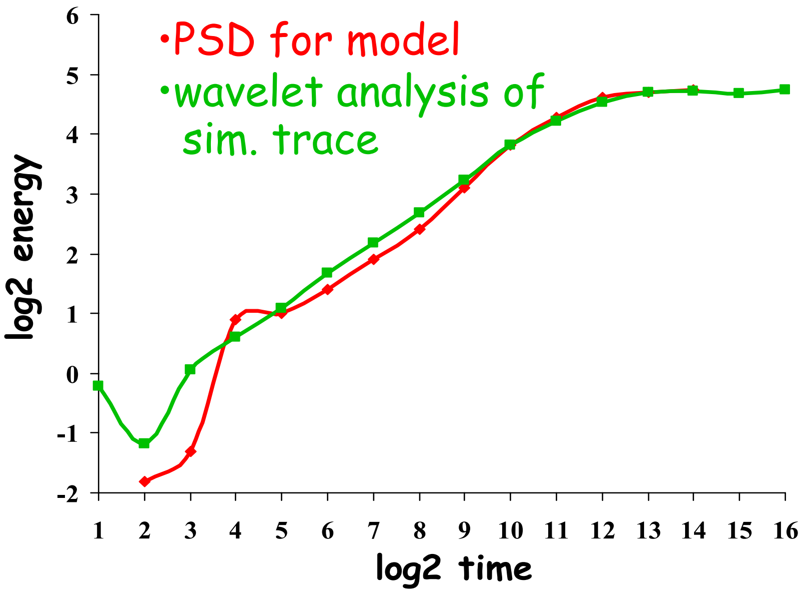
same data set size as BC97 study



TCP and long range dependence

- focus on single flow
- developed Markov chain
  - congestion avoidance (CA)
  - timeouts (TO)
- CA dominates correlation at low losses
- TO dominates correlation at high losses
- model predicts
  - no long range dependence
  - validated against simulation

from Veres, Boda Infocom00
loss rate 18%



Other work

- account for ACK loss
- sensitivity analysis of fluid models
- comparison of rate- and window-based control
- graph evolution model for Internet

Future plans

- develop error analysis for time stepped simulation
- validate ODE, fixed point models against measurements from Utah testbed
- transition technology to Nortel Networks
- QoS
  - excess bandwidth allocation
  - mix of UDP and TCP flows
- wireless